L_1 ERGODIC BEHAVIOR OF NON-NEGATIVE KERNELS

BY

DEAN L. ISAACSON AND RICHARD W. MADSEN

ABSTRACT

Some results that have been obtained in thestudy of strongly and weakly ergodic behavior of non-homogeneous stochastic kernels are generalized to the case of non-negative kernels. The first generalization simply involves extending the definitions of weakly and strongly ergiodc behavior to the case of non-negative kernels and using the ergodic coefficient which was first defined for stochastic kernels by Dobrushin and extended to non-negative kernels by Blum and Reichaw. It happens that this straightforward extension excludes many cases of non-negative kernels which do exhibit a type of ergodic behavior. In order to study these cases a definition of L_1 weakly and strongly ergodic behavior is given in which normalizing by constants is allowed. Sufficient conditions for these types of ergodic behavior are given.

1.

In this paper we consider the ergodic behavior of sequences of non-negative kernels and seek to generalize some results that have been obtained in the study of strongly and weakly ergodic behavior of non-homogeneous Markov chains.

Let $(S,\mathscr{B},\mu)$ be a σ-finite measure space and let $\{M_n(x,y)\}$ be a sequence of non-negative, measurable functions defined on $S \times S$. We consider only those sequences for which superpositions $M_{m,m+n}(x,y)$ defined by

$$M_{m,m+n}(x,y) = \int_S \cdots \int_S M_m(x,z_1)M_{m+1}(z_1,z_2)\cdots M_{m+n}(z_n,y)\mu(dz_1)\cdots\mu(dz_n)$$

exist for all m and n. Such functions $M(x,y)$ will be referred to as non-negative kernels. If it is also true that $\int_S(M(x,y)\mu(dy) = 1$ for all x, we call the kernel stochastic.

Received January 12, 1973 and in revised form January 2, 1974

In addition to the assumption of the existence of superpositions of these kernels, we assume that the kernels are sufficiently well behaved to assure that for all finite measures ζ on $(S,\mathscr{B})$ the function $g(y)$ defined by

$$g(y) = \int M_n(x,y)\zeta(dx)$$

is integrable for all n. (Henceforth, integration will be assumed over S unless otherwise indicated.)

DEFINITION 1.1. If ζ_0 is a finite measure on $(S,\mathscr{B})$, then ζ_0 will be called a starting distribution or starting measure. If, in addition, $\zeta_0(S) = 1$, then ζ_0 will be called a starting probability distribution. If ζ_0 is absolutely continuous with respect to μ, then $f_0(x) = d\zeta_0(x)/d\mu(x)$ will be called the starting density. If, in addition, $\int f_0(x)\mu(dx) = 1$, then $f_0(x)$ will be called a starting probability density.

For $\zeta_0(x)$ and $\eta_0(x)$ starting distributions, define $f_{m,n}(y) = \int M_{m,n}(x,y)\zeta_0(dx)$ and $g_{m,n}(y) = \int M_{m,n}(x,y)\eta_0(dx)$.

DEFINITION 1.2. A sequence of non-negative kernels $\{M_n\}$ will be called weakly ergodic if for all m,

$$\sup_{\zeta_0,\eta_0} \int |f_{m,n}(y) - g_{m,n}(y)|\,\mu(dy) \to 0 \text{ as } n \to \infty. \tag{1.1}$$

DEFINITION 1.3. A sequence of non-negative kernels $\{M_n\}$ will be called strongly ergodic if there exists a function $q(y)$ such that for all m,

$$\sup_{\zeta_0} \int |f_{m,n}(y) - q(y)|\,\mu(dy) \to 0 \quad \text{as } n \to \infty.$$

In studying the ergodic behavior of non-negative kernels it is useful to introduce the ergodic coefficient which was first defined for stochastic kernels by Dobrushin [3] and extended to non-negative kernels by Blum and Reichaw [1].

DEFINITION 1.4. If $M(x,y)$ is a non-negative kernel, the ergodic coefficient $\delta(M)$ is defined to be

$$\delta(M) = \sup_{x,z} \int [M(x,y) - M(z,y)]^+\mu(dy). \tag{1.2}$$

For the case of stochastic kernels $P_n(x,y)$ where all starting distributions are taken to be starting probability distributions, it is known [6] that $\{P_n\}$ is weakly ergodic if and only if $\delta(P_{m,n}) \to 0$ as $n \to \infty$ for all m. In extending this result to non-negative kernels, the convention of requiring all starting distributions to be starting probability distributions will be retained. In this case we obtain the following theorem.

THEOREM 1.1. *Let* $\{M_n\}$ *be a sequence of non-negative kernels.* $\delta(M_{m,n}) \to 0$ *as* $n \to \infty$ *for all* m *if and only if*

$$\int_E \int_S M_{m,n}(x,y)\,[\zeta_0(dx) - \eta_0(dx)] \to 0 \tag{1.3}$$

uniformly in E, ζ_0, *and* η_0 *as* $n \to \infty$ *for all* m.

PROOF. Assume $\delta(M_{m,n}) \to 0$ as $n \to \infty$ for all m. Using Inequality I of [1] with $\alpha = \zeta_0(S) - \eta_0(S) = 0$ we obtain

$$\int_E \int_S M_{m,n}(x,y)\,[\zeta_0(dx) - \eta_0(dx)] \leqq \delta(M_{m,n})\,[\zeta_0 - \eta_0]^+(S) \leqq \delta(M_{m,n}).$$

The right-hand side goes to zero uniformly in ζ_0, η_0, and E as $n \to \infty$ for all m.

Conversely, assume $\int_E \int_S M_{m,n}(x,y)\,[\zeta_0(dx) - \eta_0(dx)]\mu(dy) \to 0$ uniformly in E, ζ_0, and η_0 as $n \to \infty$ for all m. Let ζ_0 be the starting probability distribution which assigns probability measure one to the point x_1. Let η_0 be the starting probability distribution which assigns probability measure one to the point x_2. Let $E = \{y\colon \int M_{m,n}(x,y)\,[\zeta_0(dx) - \eta_0(dx)] > 0\}$. For this choice of ζ_0, η_0, and E it follows that

$$\int [M_{m,n}(x_1,y) - M_{m,n}(x_2,y)]^+ \mu(dy) \to 0 \quad \text{for all } m.$$

However, since this holds uniformly in ζ_0, η_0, and E we obtain $\sup_{x_1,x_2} \int [M_{m,n}(x_1,y) - M_{m,n}(x_2,y)]^+ \mu(dy) \to 0$ as $n \to \infty$ for all m so $\delta(M_{m,n}) \to 0$ as $n \to \infty$ for all m. ∎

REMARK. Theorem 1.1 gives conditions for weak ergodicity since (1.1) is equivalent to (1.3).

REMARK. Inequality I in [1] is stated for starting distributions that are absolutely continuous with respect to μ. The proof of the same inequality for finite signed measures as used above is straightforward.

The Blum and Reichaw inequalities are also used in [1] to obtain some sufficient conditions for strong ergodicity of non-negative kernels. However, for non-negative kernels one should not stop here, since there are many well-behaved, non-negative kernels that do not satisfy these restrictive definitions of weak and strong ergodicity. Consider, for example, $M_m = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ which yields a homogeneous chain such that if $\zeta_0 = (\frac{1}{2}, \frac{1}{2})$, then $\zeta_0 M_1 M_2 \cdots M_n = f_{1,n} \to (\infty, \infty)$ as $n \to \infty$. It is easy to see that if some norming is used at each stage, the sequence becomes ergodic. In fact, if $f_{m,n}(j)$ denotes the jth component of $f_{m,n}$ consider $f^*_{m,n}(j) = f_{m,n}(j)/(f_{m,n}(1) + f_{m,n}(2))$, $j = 1,2$. Then, for the above example, $f^*_{m,n} \to (\frac{1}{2}, \frac{1}{2})$ and $\sum_{j=1}^{2} |f^*_{m,n}(j) - g^*_{m,n}(j)| \to 0$ as $n \to \infty$ for all m and for all starting probability measures ζ_0 and η_0. This leads to a revised definition of weak and strong ergodicity which will use norming to keep $f^*_{m,n}$ bounded away from zero and infinity. Since the normalizing factor may depend on ζ_0, m, and n, a more general approach is to divide by a constant $k(\zeta_0, m, n)$. Define $f^*_{m,n}(y) = f_{m,n}(y)/k(\zeta_0, m, n)$ and $g^*_{m,n}(y) = g_{m,n}(y)/k(\eta_0, m, n)$.

We note that Conn [2] gave conditions for $|(f_{m,n}(y)/\int f_{m,n}) - (g_{m,n}(y)/\int g_{m,n})|$ to go to zero. This kind of behavior will be called pointwise weakly ergodic. Conn also found conditions for pointwise strongly ergodic behavior. To distinguish this and the stochastic case from the ergodic behavior described here, we use the terms normalized L_1 weakly ergodic (NL_1WE) and normalized L_1 strongly ergodic (NL_1SE).

DEFINITION 1.5. Let $\{M_n(x,y)\}$ be a sequence of non-negative kernels and let $\mathscr{M}$ be the family of all starting probability measures ζ. The sequence $\{M_n\}$ will be called NL_1WE if for each pair $\zeta_0, \eta_0 \in \mathscr{M}$ there exist sequences of positive constants $k(\zeta_0, m, n)$ and $k(\eta_0, m, n)$ such that for all m,

$$\sup_{\zeta_0, \eta_0 \in \mathscr{M}} \int |f^*_{m,n}(y) - g^*_{m,n}(y)| \mu(dy) \to 0 \quad \text{as } n \to \infty$$

while $\int f^*_{m,n}(y)\mu(dy) \not\to 0$. (This last condition is included to ensure that the norming constants are not chosen so large that both $f^*_{m,n}$ and $g^*_{m,n}$ go to zero.)

DEFINITION 1.6. A sequence of non-negative kernels $\{M_n(x,y)\}$ will be called NL_1SE if there exists a function $q(y)$ such that for every starting distribution ζ_0 there exists a sequence of positive constants $k(\zeta_0, m, n)$ such that

$$\sup_{\zeta_0 \in \mathscr{M}} \int |f^*_{m,n}(y) - q(y)| \mu(dy) \to 0 \quad \text{as } n \to \infty$$

for all m.

In order to use results known for ergodic behavior of stochastic kernels, the non-negative kernels will be transformed into stochastic kernels. Eigenfunctions, if they exist, can be used for making such a transformation.

Let $\Phi_n(x)$ be a positive right eigenfunction of $M_n(x, y)$ corresponding to the eigenvalue $\lambda_n > 0$. Then

$$P_n(x, y) = M_n(x, y)\Phi_n(y)/\lambda_n\Phi_n(x) \tag{1.4}$$

is a stochastic kernel. This transformation, which was used by Harris [4], will be used in the next two sections.

2. Normalized L_1 weak ergodicity

The following lemmas will be useful in the proof of the main theorem of this section.

LEMMA 2.1. *Let $g(x)$ be an integrable function and let $h(x)$ satisfy $0 \leqq 1 - \varepsilon \leqq h(x) \leqq 1 + \varepsilon$. Then*

$$\left| \int g(x)h(x)\mu(dx) \right| \leqq \left| \int g(x)\mu(dx) \right| + \varepsilon \int |g(x)| \mu(dx).$$

PROOF. Writing $g(x) = g^+(x) - g^-(x)$ and $|g(x)| = g^+(x) + g^-(x)$ and using the bounds on $h(x)$, the proof is straightforward. ▮

For further reference, we state the following properties.

Property I. There exist right eigenfunctions $\Phi_n(x)$ corresponding to $\lambda_n > 0$ such that:

i $\quad 0 < b \leqq \Phi_n(x) \leqq B < \infty$;

ii $\quad r_n = \sup_x \left| \dfrac{\Phi_n(x)}{\Phi_{n-1}(x)} - 1 \right|$ satisfies $\sum_{n=2}^{\infty} r_n < \infty$.

Property II. The sequence of stochastic kernels $\{P_n(x, y)\}$ defined by (1.4) is weakly ergodic.

Property III. The sequence of stochastic kernels $\{P_n(x, y)\}$ defined by (1.4) is strongly ergodic.

LEMMA 2.2. *Let $\{M_n(x, y)\}$ be a sequence of non-negative kernels satisfying Properties* I *and* II. *Then, given $\varepsilon > 0$ and starting distributions ζ_0 and η_0, there are sequences of normalizing constants $\{d_n(\varepsilon)\}$ and $\{e_n(\varepsilon)\}$ such that for $n \geqq N(\varepsilon)$,*

$$\int | f^*_{1,n}(y) - g^*_{1,n}(y) | \mu(dy) < \varepsilon.$$

PROOF. Define $f_n(y) = f_{1,n}(y)$. Since $\sum_{n=2}^{\infty} r_n < \infty$, it follows that $\prod_{n=2}^{\infty}(1+r_n)$ converges. Hence, given γ such that $0<\gamma<1$, there exists $N_1 = N_1(\gamma)$ such that for all $M > N_1 + 1$,

$$\prod_{n=N_1+1}^{M}(1+r_n) \leqq \prod_{n=N_1+1}^{\infty}(1+r_n) < 1+\gamma. \tag{2.1}$$

Furthermore,

$$\prod_{n=N_1+1}^{M}(1-r_n) \geqq \prod_{n=N_1+1}^{\infty}(1-r_n) > 1-\gamma. \tag{2.2}$$

Also, since by Property II $\{P_n\}$ is weakly ergodic, given m and $\gamma > 0$, there exists an $N_2 = N_2(\gamma, m)$ such that for $n \geqq N_2$,

$$\delta(P_{m,n}) < \gamma. \tag{2.3}$$

Let $\gamma = \varepsilon b/4$, $m = N_1(\gamma)$, and $N(\varepsilon) = N_2(\gamma, m)$. If ζ_0 and η_0 are the given starting distributions, define $\{d_n\}$ as follows:

$$d_n = \begin{cases} 1 & \text{if } n < m \\ \Lambda(m,n)\int f_{m-1}(y)\Phi_m(y)\mu(dy) & \text{if } n \geqq m \end{cases} \tag{2.4}$$

where $\Lambda(j,k) = \prod_{i=j}^{k}\lambda_i$. Define $\{e_n\}$ similarly using g_{m-1} instead of f_{m-1}.

Now consider $f_n(y)$ for $n \geqq N(\varepsilon) = N_2(\gamma, m)$. Since $n > m$, we can write

$$f_n(y) = \int\cdots\int f_{m-1}(z_m)M_m(z_m, z_{m+1})\cdots M_n(z_n, y)\mu(dz_m)\cdots\mu(dz_n).$$

Using (1.4), this becomes

$$\begin{aligned} f_n(y) &= \int\cdots\int f_{m-1}(z_m)\left[\lambda_m\Phi_m(z_m)P_m(z_m, z_{m+1})/\Phi_m(z_{m+1})\right]\cdots \\ &\qquad [\lambda_n\Phi_n(z_n)P_n(z_n, y)/\Phi_n(y)]\mu(dz_m)\cdots\mu(dz_n) \\ &= \int\cdots\int f_{m-1}(z_m)\Phi_m(z_m)\Lambda(m,n)P_m(z_m, z_{m+1})\cdots P_n(z_n, y) \\ &\qquad \{\phi(z_{m+1},\cdots,z_n)/\Phi_n(y)\}\mu(dz_m)\cdots\mu(dz_n) \end{aligned}$$

where $\phi(z_{m+1},\cdots,z_n) = \prod_{j=m+1}^{n}\Phi_j(z_j)/\Phi_{j-1}(z_j)$. If we define

$$f_{m-1}^{**}(y) = f_{m-1}(y)\Phi_m(y)/\int f_{m-1}(y)\Phi_m(y)\mu(dy) \tag{2.5}$$

and if we use d_n as defined in (2.4), then

$$(2.6)\qquad f_n^*(y) = f_n(y)/d_n = \int \cdots \int f_{m-1}^{**}(z_m)P_m(z_m, z_{m+1}) \cdots P_n(z_n, y)\{\phi(z_{m+1}, \cdots, z_n)/\Phi_n(y)\}\mu(dz_m) \cdots \mu(dz_n).$$

A similar expression can be given for $g_n^*(y)$. Hence

$$\int |f_n^*(y) - g_n^*(y)|\mu(dy) = \int \left| \int \cdots \int h(z_m \cdots z_n, y)\phi(z_{m+1}, \cdots, z_n)\mu(dz_m) \cdots \mu(dz_n) \right| \frac{1}{\Phi_n(y)}\mu(dy)$$

where $h(z_m, \cdots, z_n, y) = [f_{m-1}^{**}(z_m) - g_{m-1}^{**}(z_m)]P_m(z_m, z_{m+1}) \cdots P_n(z_n, y)$.

It follows from (2.1) and (2.2) that $1 - \gamma \leqq \phi \leqq 1 + \gamma$, hence (2.1) can be applied to give

$$(2.7)\qquad \int |f_n^*(y) - g_n^*(y)|\mu(dy) \leqq \int \left| \int [f_{m-1}^{**}(z_m) - g_{m-1}^{**}(z_m)]P_{m,n}(z_m, y)\mu(dz_m) \right| \frac{1}{\Phi_n(y)}\mu(dy) + \gamma \iint |f_{m-1}^{**}(z_m) - g_{m-1}^{**}(z_m)| P_{m,n}(z_m, y) \frac{1}{\Phi_n(y)}\mu(dz_m)\mu(dy).$$

Since f_{m-1}^{**} and g_{m-1}^{**} were constructed to be starting probability densities, we can construct a stochastic kernel by choosing any set $A \in \mathscr{B}$ with $0 < \mu(A) < \mu(S)$ and defining:

$$P_{m-1}^{**}(x, y) = \begin{cases} f_{m-1}^{**}(y) & \text{if } x \in A \\ g_{m-1}^{**}(y) & \text{if } x \notin A. \end{cases}$$

Define $P_{m-1}^{**}P_{m,n}(x, y)$ to be $\int P_{m-1}^{**}(x, z)P_{m,n}(z, y)\mu(dz)$, so if $x \in A$ and $z \notin A$, the first term of (2.7) becomes

$$(2.8)\qquad \int |P_{m-1}^{**}P_{m,n}(x, y) - P_{m-1}^{**}P_{m,n}(z, y)| \frac{1}{\Phi_n(y)}\mu(dy) \leqq \frac{1}{b}\int |P_{m-1}^{**}P_{m,n}(x, y) - P_{m-1}^{**}P_{m,n}(z, y)|\mu(dy) = \frac{2}{b}\delta(P_{m-1}^{**}P_{m,n}) \leqq \frac{2}{b}\delta(P_{m,n}) \leqq \frac{2\gamma}{b}$$

where the inequalities of (2.8) follow from the bounds on Φ_n, properties of the ergodic coefficient (see [5]), and inequality (2.3).

Now consider the second term of (2.7). It is less than or equal to

$$\frac{\gamma}{b}\int\int|f^{**}_{m-1}(z)-g^{**}_{m-1}(z)|P_{m,n}(z,y)\mu(dz)\mu(dy)\leqq\frac{2\gamma}{b}. \tag{2.9}$$

Combining (2.8) and (2.9) yields

$$\int|f^*_n(y)-g^*_n(y)|\mu(dy)\leqq\frac{2\gamma}{b}+\frac{2\gamma}{b}=\varepsilon.$$

It remains to show that d_n is positive. If $\Phi(x)$ is an eigenfunction of $M(x,y)$ and $0<b\leqq\Phi(x)\leqq B<\infty$, then

$$\lambda b\leqq\lambda\Phi(x)=\int M(x,y)\Phi(y)\mu(dy)\leqq B\int M(x,y)\mu(dy). \tag{2.10}$$

Since λ is positive, using equation (2.10) we see that $\int M(x,y)\mu(dy)\geqq\lambda b/B>0$, from whence it follows that

$$\begin{aligned}\int f_n(y)\mu(dy) &= \int\int f_{n-1}(x)M_n(x,y)\mu(dx)\mu(dy)\\ &\geqq \frac{\lambda_n b}{B}\int f_{n-1}(x)\mu(dx)\\ &\geqq \Lambda(1,n)(b/B)^n\zeta_0(S)>0.\end{aligned}$$

Hence d_n (and likewise e_n) must be positive. ∎

Note that in the proof, the choice of $N(\varepsilon)$ did not depend on either ζ_0 or η_0.

THEOREM 2.1. *If $\{M_n(x,y)\}$ is a sequence of non-negative kernels satisfying properties* I *and* II, *then $\{M_n\}$ is normalized L_1 weakly ergodic.*

PROOF. Let ζ_0 and η_0 be any starting distributions. It suffices to consider the case $m=1$, since for any other m the arguments are identical.

Let $\{\varepsilon_i\}$ be a sequence of constants decreasing to zero. By Lemma 2.2, for each i there exist sequences of constants $\{d_n(\varepsilon_i)\}$ and $\{e_n(\varepsilon_i)\}$ such that $n\geqq N(\varepsilon_i)\Rightarrow\int|f^*_n(y)-g^*_n(y)|\mu(dy)<\varepsilon_i$. Without loss of generality, assume that $\{N(\varepsilon_i)\}$ forms an increasing sequence. Since $\{N(\varepsilon_i)\}$ do not depend on ζ_0 or η_0, define

$$k(\zeta_0,1,n)=\begin{cases}d_n(\varepsilon_1) & n\leqq N(\varepsilon_2)\\ d_n(\varepsilon_i) & N(\varepsilon_i)<n\leqq N(\varepsilon_{i+1})\end{cases}$$

and define $k(\eta_0,1,n)$ similarly using $\{e_n(\varepsilon_i)\}$. These sequences of constants can be used to show that $\{M_n\}$ is NL_1WE directly from Definition 1.5. If $\varepsilon>0$ is given, there is some i such that $\varepsilon_i<\varepsilon$, and for any $n>N(\varepsilon_i)$, $\int|f^*_n(y)-g^*_n(y)|\mu(dy)<\varepsilon_i<\varepsilon$ independently of the choice of ζ_0 and η_0.

It remains to show that $\int f_n^*(y)\mu(dy) \nrightarrow 0$. From (2.6) we have the following inequalities where $m = N_1(\gamma)$.

$$\int f_n^*(y)\mu(dy) = \int \cdots \int f_{m-1}^{**}(z_m)P_m(z_m, z_{m+1}) \cdots P_n(z_n, y)\phi(z_{m+1}, \cdots, z_n)/\Phi(y) \quad \mu(dz_m) \cdots \mu(dz_n)\mu(dy)$$

$$\geqq \frac{1}{B} \prod_{j=m+1}^{n} (1-r_j) \int \cdots \int f_{m-1}^{**}(z_m)P_m(z_m, z_{m+1}) \cdots P_n(z_n, y)\mu(dz_m) \cdots \mu(dz_n)\mu(dy)$$

$$= \frac{1}{B} \prod_{j=m+1}^{n} (1-r_j) \geqq \frac{1}{B} \prod_{j=m+1}^{\infty} (1-r_j) \geqq \frac{1-\gamma}{B} > 0 \quad \text{for all } n. \quad ∎$$

REMARK. Since eigenfunctions satisfying Property I(i) do not always exist, we note that another means of transforming the non-negative kernels into stochastic kernels can be used.

Property I′. There exist functions $G_n(y)$ such that:

i $0 < b \leqq G_n(y) \leqq B < \infty$;

ii $I_n(x) \equiv \int M_n(x, y)G_n(y)\mu(dy)$ exists and is positive for all x;

iii there exists a sequence of constants $\{\rho_n\}$ such that

$$r_n = \sup_x \left| \frac{I_n(x)}{\rho_n G_{n-1}(x)} - 1 \right| \text{ satisfies } \sum_{n=2}^{\infty} r_n < \infty.$$

Property II′. The sequence of stochastic kernels defined by

$$R_n(x, y) = M_n(x, y)G_n(y)/I_n(x)$$

is weakly ergodic.

Note that if $\{G_n(y)\}$ are right eigenfunctions and if ρ_n is taken to be λ_n, then Properties I′ and II′ reduce to Properties I and II.

THEOREM 1.2. *If $\{M_n(x, y)\}$ is a sequence of non-negative kernels satisfying Properties I′ and II′, then $\{M_n\}$ is* $\mathrm{NL_1WE}$.

PROOF. The proof of this theorem is similar to that of Theorem 2.1.

3. Normalized L_1 strong ergodicity

We begin with two lemmas which will be useful in proving the main theorem of this section.

LEMMA 3.1. *If Property* I(ii) *is satisfied, then given $\varepsilon > 0$ there is some $N = N(\varepsilon)$ such that for all $n > N$,* $\sup|\phi(z_N, \cdots, z_n) - 1| < \varepsilon$, *where the supremum is taken over all points, $(z_N, z_{N+1}, \cdots, z_n)$, in $\times_{i=N}^{n} S$.*

PROOF. Straightforward and left to the reader.

LEMMA 3.2. *Let* $\{M_n(x,y)\}$ *be a sequence of non-negative kernels satisfying Properties* I *and* III. *Then there is a function* $q(y)$ *such that, given* $\varepsilon > 0$ *and a starting distribution* ζ_0, *there is a sequence of normalizing constants* $\{d_n(\varepsilon)\}$ *such that for* $n \geqq N(\varepsilon)$, $\int |f_n^*(y) - q(y)| \mu(dy) < \varepsilon$.

PROOF. It follows from Property I that the sequence $\{\Phi_n(y)\}$ is a Cauchy sequence and hence has a limit, say $\Phi(y)$. In fact $\Phi_n(y) \to \Phi(y)$ uniformly. From Property III, $\{P_n(x,y)\}$ is strongly ergodic, hence there exists a constant kernel $Q(x,y) = Q(y)$ to which $P_{m,n}(x,y)$ converges for all m. Now define $q(y) = Q(y)/\Phi(y)$. Write

$$f_n(y) = \int f_{m-1}(z_m) M_{m,n}(z_m, y) \mu(dz_m)$$

where $m = N(b\varepsilon/3)$ is determined as in Lemma 3.1. Define $\{d_n\}$ as in (2.4) and $f_{m-1}^{**}(y)$ as in (2.5). Then for $n > m$,

$$f_n^*(y) = f_n(y)/d_n =$$

$$\int \cdots \int f_{m-1}^{**} f(z_m) P_m(z_m, z_{m+1}) \cdots P_n(z_n, y) \{\phi(z_{m+1}, \cdots, z_n)/\Phi_n(y)\} \mu(dz_m) \cdots \mu(dz_n).$$

Hence we can write

$$\begin{aligned} &\int | f_n^*(y) - q(y) | \mu(dy) \\ &\quad = \int | f_n^*(y) - Q(y)/\Phi(y) | \mu(dy) \\ &\quad \leqq \int \Big| \int \cdots \int f_{m-1}^{**}(z_m) P_m(z_m, z_{m+1}) \cdots P_n(z_n, y) \{[\phi(z_{m+1}, \cdots, z_n) - 1]/\Phi_n(y)\} \\ &\qquad\qquad \cdot \mu(dz_m) \cdots \mu(dz_n) \Big| \mu(dy) \\ &\quad + \int \Big| \int \cdots \int f_{m-1}^{**}(z_m) P_m(z_m, z_{m+1}) \cdots P_n(z_n, y)/\Phi(y) \mu_n(dz_m) \cdots \mu(dz_n) \\ &\qquad\qquad - Q(y)/\Phi_n(y) \Big| \mu(dy) \\ &\quad + \int | Q(y)/\Phi_n(y) - Q(y)/\Phi(y) | \mu(dy). \end{aligned} \tag{3.1}$$

Consider each term of (3.1). The first term is less than or equal to

$$\int \cdots \int f^{**}_{m-1}(z_m)P_m(z_m, z_{m+1}) \cdots P_n(z_n, y) \left| \phi(z_{m+1}, \cdots, z_n) - 1 \right| / \Phi_n(y)\mu(dz_m) \cdots \mu(dz_n)\mu(dy)$$

$$\leqq \{\sup \left| \phi(z_{m+1}, \cdots, z_n) - 1 \right| / b\} \int \cdots \int f^{**}_{m-1}(z_m)P_m(z_m, z_{m+1}) \cdots P_n(z_n, y)\mu(dz_n) \cdots \mu(dz_m)\mu(dy)$$

$$= \sup \left| \phi(z_{m+1}, \cdots, z_n) - 1 \right| / b \leqq b\varepsilon/3 \cdot b = \varepsilon/3,$$

where the last inequality follows by the choice of m. Next define $P^{**}_{m-1}(x, y) = f^{**}_{m-1}(y)$ for all x. Then the second term of (3.1) can be written

$$\int \left| P^{**}_{m-1}P_{m,n}(x, y) - Q(y) \right| / \Phi_n(y)\mu(dy) \leqq \int \left| P^{**}_{m-1}P_{m,n}(x, y) - Q(y) \right| \mu(dy)/b$$

$$= \| P^{**}_{m-1}P_{m,n} - Q \| / b \leqq \| P_{m,n} - Q \| / b,$$

where the norm is that which is used in [5]. The last inequality is easy to verify using the properties of this norm and the fact that Q is a constant kernel. By strong ergodicity of $\{P_n\}$, we have $\| P_{m,n} - Q \| / b \leqq \varepsilon/3$ for $n \geqq N_1(m, b, \varepsilon)$.

Finally, since $\Phi_n(y)$ converges uniformly to $\Phi(y)$, if $n > N_2(b, \varepsilon)$ then $|\Phi_n(y) - \Phi(y)| < b^2\varepsilon/3$. In this case,

$$\int Q(y) \left| \frac{1}{\Phi_n(y)} - \frac{1}{\Phi(y)} \right| \mu(dy) = \int Q(y) \left| \Phi(y) - \Phi_n(y) \right| / \Phi(y)\Phi_n(y)\mu(dy)$$

$$\leqq \varepsilon/3 \int Q(y)\mu(dy) = \varepsilon/3.$$

Hence for $n \geqq \max(N_1, N_2)$, $\int |f^*_n(y) - q(y)| \mu(dy) < \varepsilon$. ∎

THEOREM 3.1. *If $\{M_n(x, y)\}$ is a sequence of non-negative kernels satisfying Properties* I *and* II, *then $\{M_n\}$ is normalized L_1 strongly ergodic.*

PROOF. The proof follows from Lemma 3.2 in the same way that the proof of Theorem 2.1 follows from Lemma 2.2.

COROLLARY 3.1. *If $\{M_n(x, y)\}$ is a sequence of non-negative kernels satisfying Property* I *and in addition, corresponding to λ_n there are non-negative integrable left eigenfunctions $\psi_n(y)$ such that*

$$\int |\psi_n(y) - \psi(y)| \mu(dy) \to 0 \tag{3.2}$$

and if the stochastic kernels $\{P_n\}$ satisfy $\delta(P_n) \leqq \beta < 1$, then the sequence $\{M_n\}$ is normalized L_1 strongly ergodic. Further $q(y) = \psi(y)$.

PROOF. It is easy to verify that if $\psi_n(y)$ is a left eigenfunction for $M_n(x,y)$ corresponding to λ_n, then $\psi_n(y)\Phi_n(y)$ is a left eigenfunction for $P_n(x,y)$ corresponding to the eigenvalue 1. Furthermore, we can show that $\int|\psi_n(y)\Phi_n(y) - \psi(y)\Phi(y)|\mu(dy) \to 0$, since

$$
\begin{aligned}
&\int|\psi_n(y)\Phi_n(y) - \psi(y)\Phi(y)|\mu(dy) \\
(3.3)\quad &\leqq \int |\psi_n(y)\Phi_n(y) - \psi(y)\Phi_n(y)|\mu(dy) + \int|\psi(y)\Phi_n(y) - \psi(y)\Phi(y)|\mu(dy) \\
&\leqq B \int|\psi_n(y) - \psi(y)|\mu(dy) + \int \psi(y)|\Phi_n(y) - \Phi(y)|\mu(dy).
\end{aligned}
$$

The first term of (3.3) goes to zero by condition (3.2). Since $\Phi_n(y)$ converges uniformly to $\Phi(y)$ and since $\psi(y)$ is integrable, the second term of (3.3) also goes to zero.

In view of the above remarks, Corollary 2.2 of Madsen and Isaacson [7] holds, hence $\{P_n\}$ must be strongly ergodic, and therefore Theorem 3.1 applies. We also know from [7, Cor. 2.2] that $P_n(x,y) \to \psi(y)\Phi(y) = Q(y)$. Hence in this case,

$$q(y) = Q(y)/\Phi(y) = \psi(y). \quad \blacksquare$$

REMARK. Property I(ii) imposed on the Φ_n sequence is extremely strong. In an attempt to justify the use of this condition we give an example which shows the non-sufficiency of a weaker assumption. In particular, this example shows that Property I(ii) can not be replaced by the weaker condition that $\sup_x|(\Phi_n(x)/\Phi_{n+1}(x)) - 1| \to 0$.

Define

$$
M_{2n+1} = \begin{bmatrix} \dfrac{1}{2n+1} & 1 + \dfrac{1}{(2n+1)^{\frac{1}{2}}} - \dfrac{1}{2n+1} - \dfrac{1}{(2n+1)^{\frac{3}{2}}} \\[2ex] 1 - \dfrac{1}{(2n+1)^{\frac{1}{2}}} & \dfrac{1}{2n+1} \end{bmatrix}
$$

$$
M_{2n} = \begin{bmatrix} \dfrac{1}{2n} & 1 - \dfrac{1}{(2n)^{\frac{1}{2}}} \\[2ex] 1 + \dfrac{1}{(2n)^{\frac{1}{2}}} - \dfrac{1}{2n} - \dfrac{1}{(2n)^{\frac{3}{2}}} & \dfrac{1}{2n} \end{bmatrix}.
$$

It can be shown that both of these matrices have an eigenvalue 1 and that the corresponding eigenvectors are $\Phi_{2n+1} = (1 + 1/(2n+1)^{\frac{1}{2}}, 1)'$ and $\Phi_{2n} = (1, 1 + 1/(2n)^{\frac{1}{2}})'$. Also note that $\sup_x |\Phi_n(x)/\Phi_{n+1}(x)) - 1| = 1/n^{\frac{1}{2}} \to 0$ as $n \to \infty$ but Property I(ii) fails.

Define $P_n(i,j) = M_n(i,j)\Phi_n(j)/\Phi_n(i)$ so

$$P_n = \begin{bmatrix} \frac{1}{n} & 1 - \frac{1}{n} \\ 1 - \frac{1}{n} & \frac{1}{n} \end{bmatrix}.$$

Now $\delta(P_n) = 1 - 2/n$ so $\delta(P_{m,n}) \leqq \prod_{k=m+1}^{n} \delta(P_k) \to 0$ and hence $\{P_n\}$ is weakly ergodic. In fact, $\psi_n = (\frac{1}{2}, \frac{1}{2})$ is the left eigenfunction for all n so $\{P_n\}$ is strongly ergodic [7].

Now consider $(1,1)M_nM_{n+1} \cdots M_{n+k}$ and $(1,1)M_nM_{n+1} \cdots M_{n+k+1}$ for n odd and k even. The first product yields a vector with a very large term in the second coordinate and a term near zero in the first coordinate. The second product yields a vector with the large term in the first coordinate. This large term, which goes to infinity as $k \to \infty$ for all n, continues to alternate between the first and second coordinate. Hence the only normalization that will reduce $(1,1)M_{n,n+k}$ to a constant vector is a normalization that brings everything to zero. Therefore $\{M_n\}$ is not strongly ergodic.

This example shows that one can not replace Property I(ii) with the weaker assumption that $\sup_x |(\Phi_n(x)/\Phi_{n+1}(x)) - 1| \to 0$ in the above theorems.

Acknowledgement

The authors would like to thank the referee for his helpful comments.

References

1. J. R. Blum and M. Reichaw, *Two integral inequalities*, Israel J. Math. 9 (1971), 20–26.

2. P. S. Conn, *Asymptotic properties of sequences of positive kernels*, Ph.D. dissertation Iowa State University, 1969.

3. R. L. Dobrushin, *Central limit theorem nonstationary for Markov chains I, II*, Theory Probability Appl. 1 (1956), 65–80; English trans., 329–383.

4. T. E. Harris, *The Theory of Branching Processes*, Springer-Verlag, Berlin, 1963.

5. R. W. Madsen, *A note on some ergodic theorems of A. Paz*, Ann. Math. Statist. 42 (1971), 405–408.

6. R. W. Madsen, *Asymptotic properties of superpositions of non-negative kernels*, Ph.D. dissertation, Iowa State University, 1971.

7. R. W. Madsen, and D. L. Isaacson, *Strongly ergodic behavior for non-stationary Markov processes*, Ann. Prob. **1** (1973), 329–335.

8. A. Paz, *Ergodic theorems for infinite probabilistic tables*, Ann. Math. Statist. **41** (1970), 539–550.

Department of Mathematics and Statistics
Iowa State University
Ames, Iowa, U.S.A.

and

Departmemt of Statistics
University of Missouri–Columbia
Columbia, Missouri, U.S.A